# Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools: reply to comment

EDGAR GUEVARA,[1,2,*] JUAN CARLOS TORRES-GALVÁN,[2] MIGUEL G. RAMÍREZ-ELÍAS,[3] CLAUDIA LUEVANO-CONTRERAS,[4] AND FRANCISCO JAVIER GONZÁLEZ[2]

[1] *CONACYT-Universidad Autónoma de San Luis Potosí, Mexico*
[2] *Terahertz Science and Technology Center (C2T2) and Science and Technology National Lab (LANCyTT), Universidad Autónoma de San Luis Potosí, Mexico*
[3] *Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, Mexico*
[4] *Department of Medical Sciences, University of Guanajuato, Leon, Mexico*
*[*] eguevara@conacyt.mx*

**Abstract:** We show the spectra of advanced glycation products in response to recent comments made by Bratchenko *et al*. Our results suggest that information retrieved by Raman spectroscopy is relevant to screening diabetic patients, however, the comparison carried out in our paper, between ANN and SVM, was not fair, because of the erroneous PCA selection procedure and different sources of variation present in the analysis.

## 1.    Reply

We thank Bratchenko *et al.* [1] for their critical comment to our paper [2]. Indeed, there was a miscalculation of the spot size, which was stated to be 200μm, but it was approximately 4mm. However, this spot size refers to the $1/e^2$ width, which leads to an effective irradiance of 1.45 W/cm$^2$, well within the ANSI standard. Furthermore, none of the patients showed signs of pain, numbness, tingling or discomfort during the Raman spectroscopy measurements and there were no visible signs of the slightest irritation on the acquisition sites.

We need to clarify that the standard deviations depicted in Fig. 1 of our previous work [2] were computed from the averaged spectra of eleven diabetic patients and nine controls respectively.

Regarding the possible overfitting of our artificial neural network (ANN), it is mentioned in our work that "all metrics of performance reported in this paper were computed by averaging the results of one thousand 10-fold cross-validation runs", this was done precisely in order to take into account large difference in variance between the test set and the training set. Moreover, our runs showed that the classifier achieved a low training error, while generalizing to the test set with an acceptable error, therefore, showing no signs of overfitting, despite the comparatively large size of the model. Finally, we would like to clarify that both the ANN and SVM results shown in Fig. 3 and 6 are the result of those 1000 reinitializations. Nonetheless, the random initializations of the ANN represent an additional source of variation to the cross-validation partitions, hence the comparison between both classifiers is not fair. It is to be noted that the ANN converged in approximately 4-5 different configurations, where most of the runs (~985/1000) presented a very closely related set of weights.

We agree with the authors that each criteria of principal component (PC) selection may provide a different number of PCs. It is to be noted that the results shown with 2 PCs (Fig. 5) were done to illustrate graphically the capabilities of discrimination using only two PCs. The results shown in Fig. 3 were obtained with all the PCs kept by Bartlett's criterion. We acknowledge that by performing PCA out of the cross-validation loop, and retaining a portion of the 19 principal
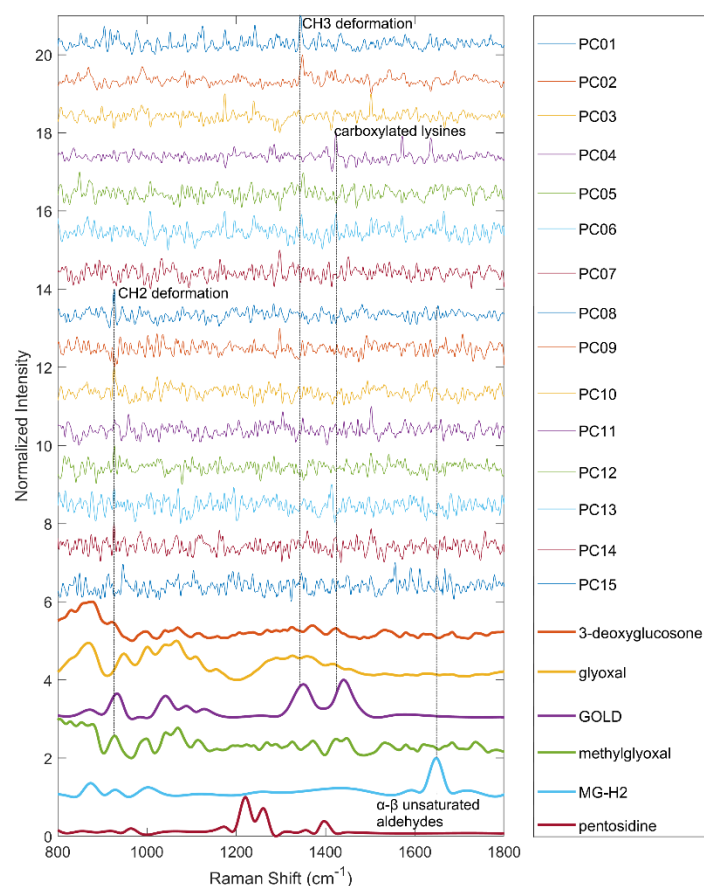
**Fig. 1.** Raman spectra of AGEs and PC loadings of diabetic patients from data acquired at the cubital vein.

components represents rather the decorrelation of this particular dataset and not the selection of important features. Indeed, the application of Bartlett's criterion was misleading in this context, because the number of components of the training set after mean-centering is 15, and all of them were retained in our paper. We admit the erroneous procedure of doing PCA on the entire data set, which meant that the training and validations subsets are also used in the training of the classifier.

In this regard, PCA was removed from the analysis and the mean spectrum from the training set was subtracted from the test and training sets at each cross-validation fold. The accuracy of SVM is compared to that of ANN, as well as that of a simple k-nearest neighbor, using different distances in the following Table 1:

**Table 1. Comparison between classifiers**

| Fold | Accuracy (%) | | | | | | | |
|------|------|------|------|------|------|------|------|------|
| | ANN | | SVM | | KNN (euclidean) | | KNN (correlation) | |
| | Train | Test | Train | Test | Train | Test | Train | Test |
| Mean | 100 | 90 | 100 | 50 | 75.5 | 50 | 74.4 | 65 |
| Std | 0 | 31.6 | 0 | 23.5 | 7.0 | 33.3 | 8.4 | 24.1 |

We show the spectra of advanced glycation products in Fig. 1 to provide the reader additional information about the shape of the PCs and the advanced glycation end products (AGEs) spectra. For instance, Furthermore, all the Raman spectra of both the patients and the AGEs can be found on the Kaggle database "Raman spectroscopy of Diabetes" by the authors [3]. The dataset consists of 20 spectra per acquisition site (80 in total), and unfortunately we do not have the unaveraged spectra because averaging was performed by the acquisition software provided with the instrument.

We must note that the correlation analysis on the AGEs spectra was intended to provide explanatory value to the implemented classifiers, but it does not mean that the PCs scores with very low correlation coefficients are useless in DM classification, it only means that they are not closely related to the Raman spectra of the aforementioned molecules. There is the possibility that the combination of other molecules present in diabetic patient may permit the classification and this could be investigated in a future work.

It has to be emphasized that our classifiers performance was compared to the performance due to random chance at a significance level of 0.05, achieving a statistically significant better performance for all the datasets using ANN and SVM, except the earlobe and inner arm dataset using SVM. Our statement "Using ANN, the skin location with the highest classification accuracy is the inner arm, with 96%" should indeed be corrected to "Using ANN, the skin location with the **largest** classification **AUC** is the inner arm, with 96%". We thank the authors for pointing out this issue.

However we must recognize that the choice of the number of hidden neurons, as per Huang et al. [4] is not the most appropriate decision, given that maximum accuracy is obtained with only three neurons, and it decreases slightly with the size of the ANN, as shown by Fig. 2. Hence, this was a misunderstanding of the model selection logic, from our part.
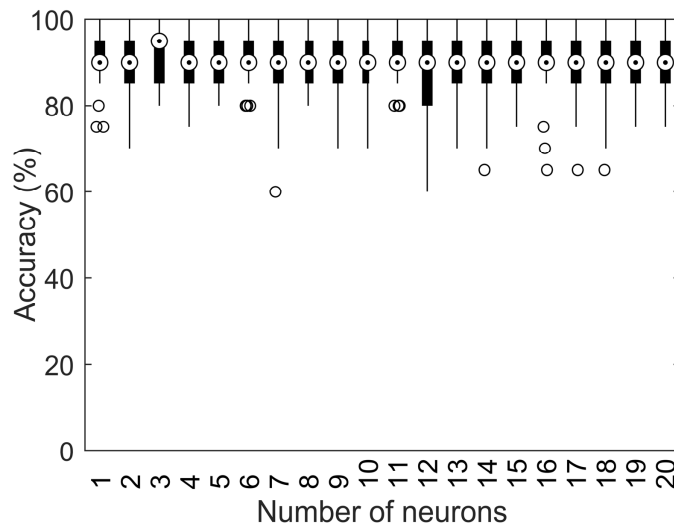


**Fig. 2.** Accuracy of the proposed ANN for different number of neurons.

We have chosen the simpler 3-neuron model and re-run the analysis over 1000 random reinitializations for the spectra acquired on the cubital vein and we report the best accuracy = 90% on a ten-fold cross-validation model, from a single random partition. The only misclassified subject was the first diabetic patient, which was incorrectly labeled as control by this model. However, visually correlating the set of weights of this optimal configuration to spectral features is difficult.

With respect to the impact of age and gender on the classification, it is very premature to assess their influence given the small sample size.

In conclusion, our results suggest that information retrieved by Raman spectroscopy is relevant to screening diabetic patients, however, the comparison between ANN and SVM was not fair, due to the fact that random initializations of the ANN represented an additional source of variation to that of the cross-validation partitions.

## Funding

## Disclosures

The authors declare that there are no conflicts of interest related to this article.

## References

1. I. A. Bratchenko, D. N. Artemyev, J. A. Khristophorova, and L. A. Shamina, "Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools": comment," Biomed. Opt. Express **10**(9), 4489–4491 (2019).
2. E. Guevara, J. C. Torres-Galván, M. G. Ramírez-Elías, C. Luevano-Contreras, and F. J. González, "Use of Raman spectroscopy to screen diabetes mellitus with machine learning tools," Biomed. Opt. Express **9**(10), 4998–5010 (2018).
3. E. Guevara, J. C. Torres-Galván, M. G. Ramírez Elías, C. Luevano-Contreras, and F. J. González, "Raman spectroscopy of Diabetes," https://kaggle.com/codina/raman-spectroscopy-of-diabetes.
4. G.-B. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," IEEE Trans. Neural Netw. **14**(2), 274–281 (2003).